



Massive Horizontal Gene Transfer in Bdelloid Rotifers

Citation

Gladyshev, Eugene A., Matthew Meselson, and Irina R. Arkhipova. 2008. Massive horizontal gene transfer in bdelloid rotifers. *Science* 320(5880): 1210-1213.

Published Version

<http://dx.doi.org/10.1126/science.1156407>

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:3120157>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

20. The authors thank S. Liebhaber for the PCBP1 plasmid, protein samples, and helpful discussions; H. Levin for helpful discussions and ferritin plasmids; P. Ponka for salicylaldehyde isonicotinoyl hydrazine; and T. Rouault for helpful discussions, cell lines, and the IRP2 antibody. These studies were supported by the Intramural Research

Program of the National Institute of Diabetes and Digestive and Kidney Diseases (H.S. and C.C.P.) and by NIH R01 DK068139 (K.Z.B. and T.L.S.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/320/5880/1207/DC1

Materials and Methods
Figs. S1 to S7
References

11 March 2008; accepted 18 April 2008
10.1126/science.1157643

Massive Horizontal Gene Transfer in Bdelloid Rotifers

Eugene A. Gladyshev,¹ Matthew Meselson,^{1,2*} Irina R. Arkhipova^{1,2*}

Horizontal gene transfer in metazoans has been documented in only a few species and is usually associated with endosymbiosis or parasitism. By contrast, in bdelloid rotifers we found many genes that appear to have originated in bacteria, fungi, and plants, concentrated in telomeric regions along with diverse mobile genetic elements. Bdelloid proximal gene-rich regions, however, appeared to lack foreign genes, thereby resembling those of model metazoan organisms. Some of the foreign genes were defective, whereas others were intact and transcribed; some of the latter contained functional spliceosomal introns. One such gene, apparently of bacterial origin, was overexpressed in *Escherichia coli* and yielded an active enzyme. The capture and functional assimilation of exogenous genes may represent an important force in bdelloid evolution.

Horizontal gene transfer (HGT), the movement of genes from one organism to another by means other than direct descent (vertical inheritance), has been documented in prokaryotes (1) and in phagocytic and parasitic

unicellular eukaryotes (2–4). Despite the large number of sequenced genomes, documented HGT is rare in metazoans, at least in part because of the sequestration of the germ line (5, 6). HGT may be facilitated by long-term association with

organelles or with intracellular endosymbionts and parasites (7, 8), or it may involve transposable elements (TEs) (9, 10).

Bdelloid rotifers are small freshwater invertebrates that apparently lack sexual reproduction and can withstand desiccation at any life stage (11, 12). Their genomes contain diverse TEs, including DNA transposons and retrovirus-like *env*-containing retrotransposons, such as *Juno* and *Vesta*, possibly acquired from exogenous sources and concentrated near telomeres (13, 14). We investigated TE distribution in bdelloids by sequencing clones from an *Adineta vaga* fosmid library hybridizing to *Juno1* probes. Unexpectedly, in two *Juno1* long terminal repeat (LTR)-containing clones (contigs Av240A and Av212A), we found 10 protein-coding sequences (CDS) yielding strong database hits (BLAST E-values of $8E^{-102}$ to 0.0) to bacterial and fungal genes (Fig. 1A, Table 1, fig. S1A, and table S1). Half of these CDS have no metazoan orthologs, and three apparently bacterial CDS are interrupted by canonical spliceosomal introns, which are nonexistent in bacteria.

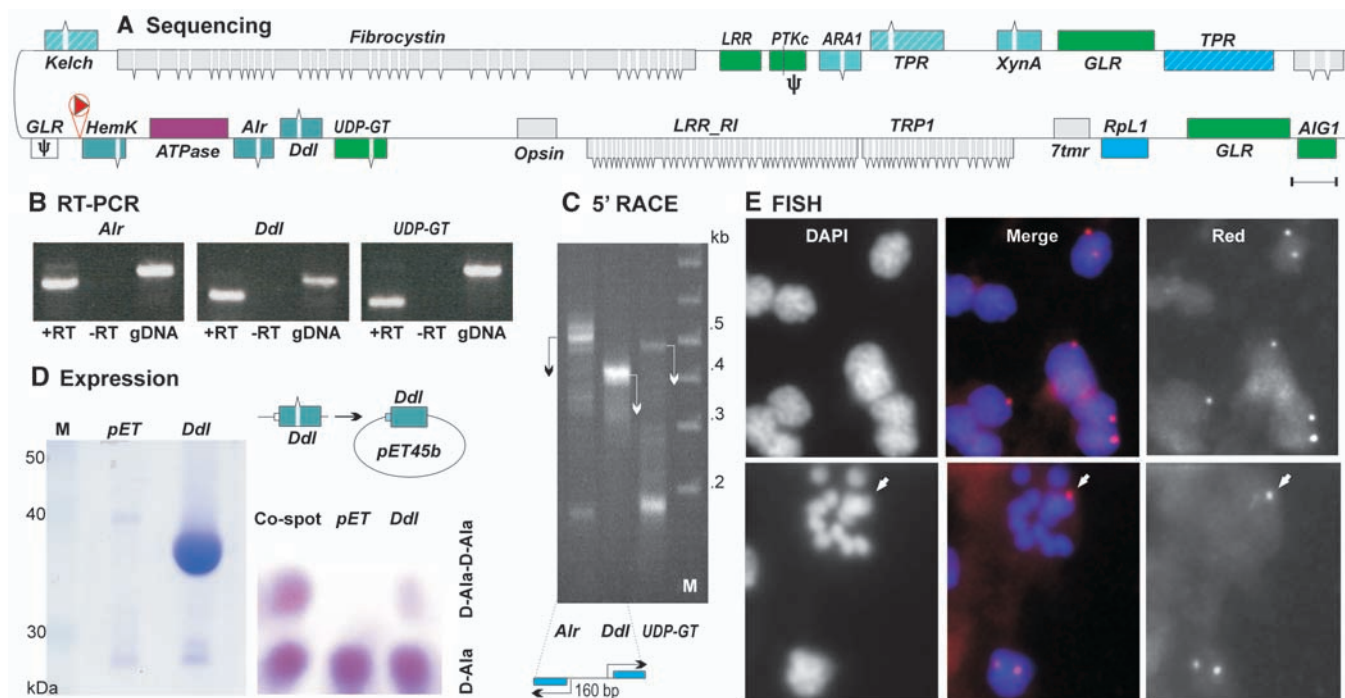


Fig. 1. Structural and functional analysis of the Av240 genomic region. (A) The 85-kb Av240B contig. The colinear 50-kb Av240A contig contains the *Juno1* LTR (red triangle) and extends from TPR to LRR_RI (not shown). CDS (boxes) are colored according to their putative origin: metazoan, gray; bacterial, blue; fungal, purple; plant, green; indeterminate, striped; hypothetical, white. Intron positions are indicated. Pseudogenes are denoted by ψ , and defects in their reading frames appear as vertical lines. Scale bar, 1 kb. (B) RT-PCR analysis of *Alr*, *Ddl*, and UDP-glycosyltransferase genes. Unspliced and spliced RNA are visible in +RT lanes; –RT, no reverse transcriptase; gDNA,

genomic DNA control. (C) 5' RACE analysis of transcription start sites (arrows) for genes in Fig. 1B. (D) Expression in *E. coli* of the *A. vaga* *Ddl* cloned in pET45b to yield a protein fused to the 6 \times His tag. Left, SDS–polyacrylamide gel electrophoresis (PAGE) analysis of the *Ddl* protein after purification by metal-affinity chromatography; right, thin-layer chromatography after incubation of purified AvDdl with D-Ala; co-spot, control for D-Ala and D-Ala-D-Ala migration; pET, vector with no insert. (E) FISH of the Av240A probe, labeled with the red fluorophore, to *A. vaga* embryo nuclei. The arrow points to a signal in a nucleus with condensed chromosomes.

Fluorescent in situ hybridization (FISH) with a probe from Av240A confirmed that these CDS reside in the *A. vaga* genome (Fig. 1E). Their hybridization pattern resembles that of known telomeric fosmids (15), suggesting proximity to telomeres. The appearance of two hybridizing sites in some nuclei is consistent with the genome structure of bdelloids in which chromosomes occur as colinear pairs (16, 17). Indeed, we found colinear partners of both Av240A and Av212A (Av240B and Av212B) with overall pairwise divergence (4%) similar to that in other regions of bdelloid genomes.

The cluster of foreign genes near the *Juno1* LTR in Av240 includes two divergently oriented genes for enzymes involved in bacterial cell wall peptidoglycan biosynthesis—*Alr*, encoding alanine racemase, and *Ddl*, encoding D-Ala-D-Ala ligase—adjacent to a uridine diphosphate (UDP)-glycosyltransferase apparently of plant origin (Fig. 1A). Reverse transcription polymerase chain reaction (RT-PCR) shows that all three genes are transcribed and that their introns are properly spliced (Fig. 1B). 5' RACE (rapid amplification of cDNA ends) demonstrates that the UDP-glycosyltransferase mRNA is trans-spliced, as are numerous bdelloid genes (18), and that *Alr* and *Ddl* transcription initiates at oppositely oriented pro-

motors located between them (Fig. 1C). Furthermore, the purified *A. vaga* AvDdl protein overexpressed in *E. coli* catalyzes the synthesis of the D-Ala-D-Ala dipeptide from D-Ala in vitro (Fig. 1D).

In addition to ubiquitous bacterial genes, such as *Alr* and *Ddl*, we identified genes occurring in only a few species of bacteria and fungi. The *XynB*-like gene (figs. S1A and S2B) apparently represents a fusion between two different conserved domains and is found in only 10 bacterial species. Next to the *Alr-Ddl* pseudo-operon, there is a *HemK*-like methyltransferase and a putative adenosine triphosphatase. These two genes are also rare: They are found in only four genera of proteobacteria and three genera of filamentous fungi. In the bacterium *Methylobacillus flagellatus* and in the fungus *Phaeosphaeria nodorum* they are adjacent and in the same relative orientation as in *A. vaga*, indicating that they might have been acquired from a single source.

We also found genes with similarity to those present in both metazoans and nonmetazoans. We characterized the foreignness of such genes with an alien index (AI), which measures by how many orders of magnitude the BLAST E-value for the best metazoan hit differs from that for the best nonmetazoan hit (Table 1 and table S1) (19). We tested the AI validity by phylogenetic analyses of all CDS with AI > 0 yielding metazoan hits, excluding those with repetitive sequences. We found that AI \geq 45, corresponding to a difference of \geq 20 orders of magnitude between the best nonmetazoan and metazoan hits, was a good

indicator of foreign origin, as judged by phylogenetic assignment to bacterial, plant, or fungal clades (Fig. 2 and fig. S2A). Genes with $0 < \text{AI} < 45$ were designated indeterminate, because their phylogenetic analysis may or may not confidently support a foreign origin. Four *FabG*-like genes for short-chain dehydrogenases/reductases (SDH), from two different SDH subfamilies, are most likely of bacterial origin (AI = 98/92/88/45; Fig. 2A). The *A. vaga* galacturonidase (AI = 212) appears to be of fungal origin (Fig. 2B), and the UDP-glycosyltransferase in Av240 (AI = 28, indeterminate) belongs to a plant clade (Fig. 2C). Two genes, *XynB* and *MviM*, had sufficient nucleotide sequence similarity (~70%) to bacterial homologs for phylogenetic reconstruction and determination of nonsynonymous and synonymous divergence (fig. S2B).

We extended our search for foreign genes to two pairs of contigs ending with telomeric repeats (telomeres K and L) (15) (fig. S1, B and C). In addition to various TEs, telomeric repeats, and *Athena* retroelements characteristic of bdelloid telomeric regions, we found additional examples of foreign genes, including a pseudogene of fungal origin (putative urea transporter; Table 1) with three frame-shifts and two in-frame stop codons in one of the two colinear homologs of telomere L. Additionally, we identified foreign genes sandwiched between short stretches of telomeric repeats, suggesting their addition to deprotected chromosome ends (fig. S1C).

We also observed examples of possible loss of genes and TEs from telomeres (fig. S3A), such

¹Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. ²Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA 02543, USA.

*To whom correspondence should be addressed. E-mail: msm@wjh.harvard.edu (M.M.); iarkhipova@mbl.edu (I.R.A.)

Table 1. Representative bdelloid CDS of foreign origin homologous to genes with known function. For a complete list and additional information on each CDS, see table S1. Data are from BLASTP similarity searches, as described in (19). Asterisks indicate putative pseudogenes.

Gene ID, name	Contig ID	Introns	AI	% Identity to best hit	Best hit, E-value	Best hit, metazoan	Best hit, taxonomy	Definition
AV10027_ <i>XynB</i>	Av212A	0	460	63	0.00E+00	No hits	Bacteria; Bacteroidetes	Xylosidase/arabinosidase
AV10001_ <i>NRPS</i>	Av110A	10	460	32	0.00E+00	No hits	Bacteria; (Proteobacteria/ Cyanobacteria)	Nonribosomal peptide synthetase
AV10134_ <i>PheA</i>	161F07	0	400	61	1.00E-174	No hits	(Fungi; Bacteria)	Monooxygenase, FAD dependent
AV10002_ <i>TrkA</i>	Av110A	0	379	54	1.00E-175	4.00E-11	Bacteria; Proteobacteria	Monooxygenase, NAD dependent
PR10002_ <i>MviM</i>	182F10	0	327	67	1.00E-149	2.00E-07	Bacteria; (Acidobacteria/ Chloroflexi)	Oxidoreductase
PR10010_ <i>DAP2</i>	182F10	0	310	27	1.00E-140	1.00E-05	Bacteria; (Acidobacteria/ Proteobacteria)	Prolyl oligopeptidase*
AV10104_ <i>Dur3</i>	AvTell.B	1	243	44	1.00E-132	4.00E-27	Eukaryota; Fungi	Urea active transporter*
PR10012_ <i>RamA</i>	182J17	0	246	31	1.00E-107	No hits	(Bacteria; Fungi)	α -L-Rhamnosidase
AV10121_ <i>NRPS</i>	9907	4	237	30	1.00E-103	No hits	Bacteria; Cyanobacteria	Nonribosomal peptide synthetase
AV10153_ <i>XghA</i>	210B3	0	212	50	1.00E-108	2.00E-16	Eukaryota; Fungi	Endo-xylagalacturonan hydrolase
AV10042_ <i>HemK</i>	Av240B	1	199	56	2.00E-91	1.00E-04	Bacteria; Proteobacteria	HemK-like methyltransferase
AV10092_ β -Gal	AvTell.A	0	153	33	1.00E-105	4.00E-39	Eukaryota; Viridiplantae	β -D-Galactosidase
AV10044_ <i>Alr</i>	Av240B	1	152	38	1.00E-67	No hits	Bacteria; Bacteroidetes	Alanine racemase
AV10025_ <i>AMH</i>	Av212A	1	150	52	8.00E-77	2.00E-11	Eukaryota; Fungi	Amidohydrolase
AV10045_ <i>Ddl</i>	Av240B	1	138	40	1.00E-60	No hits	Bacteria; Bacteroidetes	D-Alanine-D-alanine ligase
AV10140_ <i>PLDc</i>	193E18	2	126	31	1.00E-55	No hits	Eukaryota; Fungi	Phospholipase-D active site motif protein*
AV10016_ <i>FabG</i>	Av212A	0	98	58	1.00E-74	8.00E-32	Bacteria	Short-chain dehydrogenase/reductase
AV10109_ <i>FabG</i>	AvTell.B	0	92	57	4.00E-73	5.00E-33	Bacteria	Short-chain dehydrogenase/reductase*
AV10011_ <i>FabG</i>	Av212A	0	88	54	6.00E-67	2.00E-28	Bacteria	Short-chain dehydrogenase/reductase
AV10071_ <i>HAL</i>	AvTelK.A	0	77	48	2.00E-61	1.00E-27	Bacteria	Histidine ammonia-lyase
AV10095_ <i>GCN5</i>	AvTell.A	0	59	35	2.00E-27	No hits	Bacteria; Proteobacteria	GCN5-related N-acetyltransferase ^{+/} *
AV10158_ <i>FabG</i>	210B3	2	46	41	2.00E-39	2.00E-19	Bacteria	Short-chain dehydrogenase/reductase

as the metazoan long-chain acyl-coenzyme A synthetase (ACS) gene fragment on telomere K, which has an apparently intact 5' sequence but is 3'-truncated by telomeric repeat addition to exon 2 (fig. S3A). Single-telomere length analysis PCR (STELA-PCR) (20) verified its telomeric localization (fig. S3B). No colinear partner of ACS was found, and the lack of RT-PCR product suggests that it is transcriptionally inactive or that its transcript is unstable (fig. S3C).

Two other contigs containing telomeric repeats (Table 1 and fig. S1D) had apparently intact genes for two nonribosomal peptide synthetases (NRPSs), large enzymes responsible for synthesis in bacteria and fungi of biologically active peptides including antibiotics and toxins (21).

This finding suggests that bdelloid biosynthetic activity includes the production of secondary metabolites.

We also examined sequences in the vicinity of retrovirus-like elements in *Philodina roseola*, a bdelloid that separated from *A. vaga* tens of millions of years ago (22). The *P. roseola* telomere P (15) contains a *gag* gene from a retrovirus-like element similar to *A. vaga Juno1*, named *Juno2*, which is 3'-truncated by *P. roseola* telomeric repeats (fig. S3A). We probed a *P. roseola* genomic library with this *gag* fragment and found that two of five *Juno2* copies are surrounded by foreign genes (Table 1 and fig. S1E). Thus, extensive HGT probably represents a general feature of bdelloid rotifers.

We have not yet found a case of HGT in which the donor species could be identified by near-identity of a bdelloid gene to that of a putative donor, as in some other taxa (7, 8, 23), or in which the time of transfer could be estimated from the degree of synonymous difference, presumably because the transferred genes and their relatives are not yet represented in databases or because they have resided in bdelloids long enough to have substantially diverged. The similarity of the length distribution and sequence of introns in the putatively foreign genes, including genes of apparently bacterial origin, to that typical of other bdelloid genes (fig. S4) and the similarity of nucleotide composition and codon usage (table S1) suggest that some foreign genes were acquired long enough ago to conform to their host genomes, whereas other intronless genes such as *MviM* and *XynB* (fig. S2B) may have arrived more recently.

Our examination of ~1 Mb of bdelloid telomeric/TE-rich sequence (about 1% of the genome) revealed dozens of genes that are of foreign (bacterial, plant, or fungal) origin, and about twice as many indeterminate genes, some of which may be foreign (Figs. 2 and 3, Table 1, and table S1). The AI values for all the CDS in bdelloid telomeric/TE-rich regions indicate that a substantial fraction resulted from HGT, with about one-third having AI > 45 and the majority having AI > 0. In contrast, inspection of comparable DNA amounts from *P. roseola* and *A. vaga* gene-rich regions surrounding *hsp82*, histone, and Hox genes (16, 17, 19) shows that these regions contain mostly genes with known metazoan relatives and are virtually devoid of TEs (Fig. 3, A and B). Indeed, the AI distribution of CDS in these regions is indistinguishable from that for translated expressed sequence tags (ESTs) from the monogonont rotifer *Brachionus plicatilis* (24) and for CDS from *Caenorhabditis elegans* (including those near telomeres) and *Drosophila melanogaster* (Fig. 3C).

The concentration of foreign genes near telomeres may result from preferential incorporation at or near chromosome ends, perhaps at deprotected telomeres, or from selection against more proximal insertions that disrupt essential sequences or are associated with deleterious rearrangements caused by ectopic repair of double-strand breaks via capture of foreign DNA (15, 25, 26). The lack of orientation bias with respect to telomeres indicates that their acquisition does not involve an RNA intermediate copied directly into cDNA, as occurs during terminal retrotransposition. Among ORFs with AI > 45 (Table 1), we found 5 in various degrees of decay and 17 that appear to be intact. Most of the intact foreign genes code for simple enzymatic functions such as carbohydrate decomposition (Table 1) and are not parts of multicomponent pathways that might not function properly after transfer to a distant host (27).

The apparent HGT cannot reasonably be explained by vertical inheritance from the com-

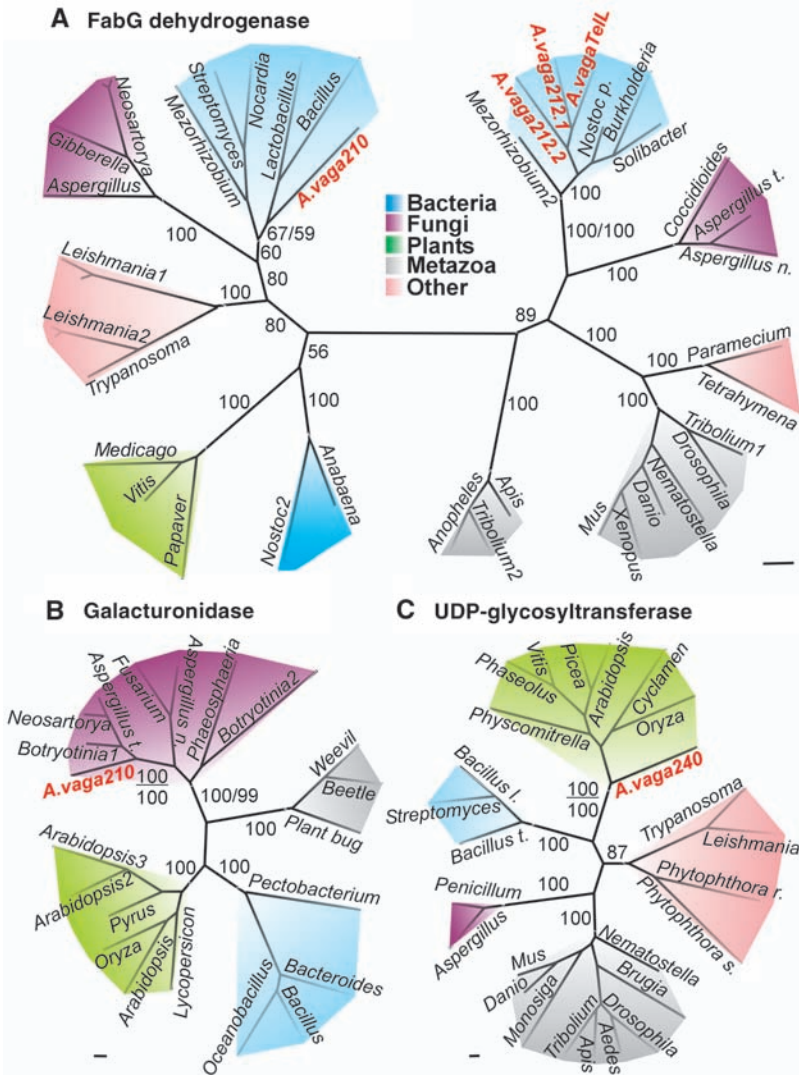
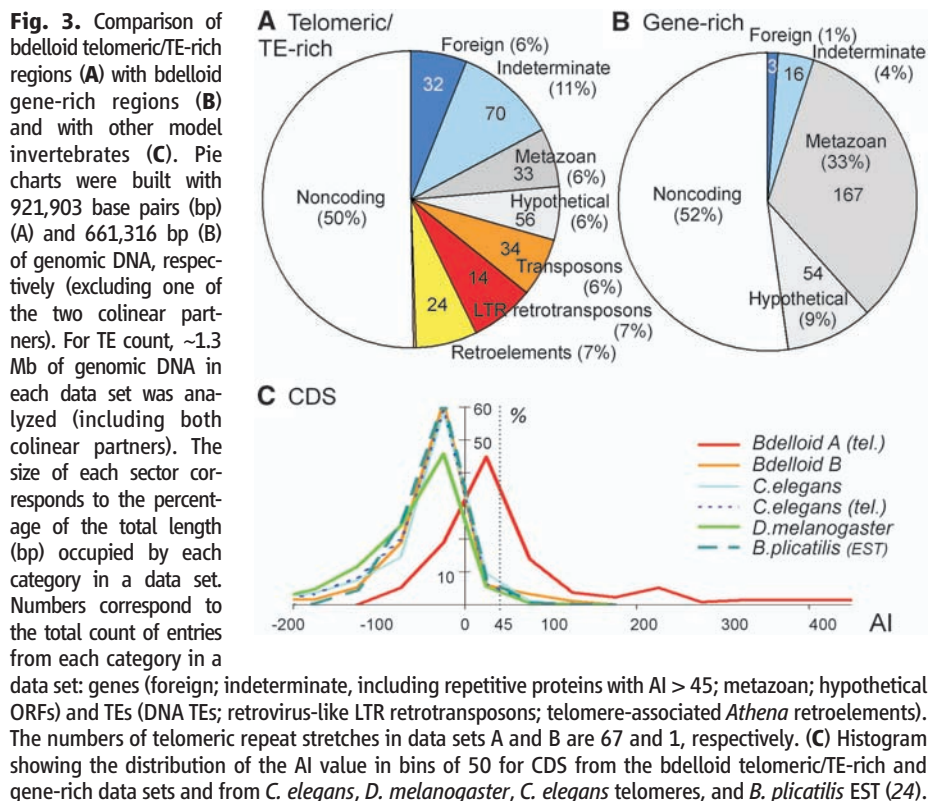


Fig. 2. Bayesian analysis of *A. vaga* CDS from three different kingdoms. Clades with different taxonomic affiliation are color-coded as in Fig. 1; CDS from *A. vaga* are in red. (A) Two subfamilies of *FabG* short-chain dehydrogenases/reductases, with four *A. vaga FabG*-like genes of bacterial origin (one contains an intron; three are intronless and could have entered independently or undergone duplication). (B) Galacturonidases and the corresponding *A. vaga* gene of fungal origin. (C) UDP-glycosyltransferases and the corresponding *A. vaga* gene of plant origin. Bayesian posterior probabilities are shown; for branches leading to *A. vaga*, neighbor-joining bootstrap support values are also indicated. For alignments, see table S11. Scale bars, 0.1 amino acid substitution per site.



mon universal ancestor and subsequent loss in other kingdoms and in at least four major metazoan branches preceding Rotifera on the evolutionary tree (28), and is inconsistent with our finding of genes representing fusions between domains of prokaryotic and eukaryotic origin. It may be that HGT is facilitated by membrane disruption and DNA fragmentation and repair associated with the repeated desiccation and recovery experienced in typical bdelloid habitats, allowing DNA in ingested or other environmental material to enter bdelloid genomes (12, 29). Whether there may also be homologous replacement by DNA segments released from related individuals remains to be seen. If there is, bdelloid rotifers may experience genetic exchange resembling that in sexual populations (30). Although the adaptive importance of such massive HGT remains to be elucidated, it is evident that such events have frequently occurred in the genomes of bdelloid rotifers, probably mediated by their unusual lifestyle.

References and Notes

- Y. Boucher *et al.*, *Annu. Rev. Genet.* **37**, 283 (2003).
- W. F. Doolittle, *Trends Genet.* **14**, 307 (1998).
- B. Loftus *et al.*, *Nature* **433**, 865 (2005).
- H. G. Morrison *et al.*, *Science* **317**, 1921 (2007).
- J. O. Andersson, *Cell. Mol. Life Sci.* **62**, 1182 (2005).
- L. A. Katz, *Int. J. Syst. Evol. Microbiol.* **52**, 1893 (2002).
- U. Berghthorsson, K. L. Adams, B. Thomason, J. D. Palmer, *Nature* **424**, 197 (2003).
- J. C. Dunning Hotopp *et al.*, *Science* **317**, 1753 (2007); published online 30 August 2007 (10.1126/science.1142490).
- M. G. Kidwell, *Annu. Rev. Genet.* **27**, 235 (1993).
- M. Syvanen, C. I. Kado, Eds., *Horizontal Gene Transfer* (Academic Press, London, 2002).
- B. B. Normark, O. Judson, N. Moran, *Biol. J. Linn. Soc.* **79**, 69 (2003).
- J. Lapinski, A. Tunncliffe, *FEBS Lett.* **553**, 387 (2003).
- I. R. Arkhipova, M. Meselson, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11781 (2005).
- E. G. Gladyshev, M. Meselson, I. R. Arkhipova, *Gene* **390**, 136 (2007).
- E. A. Gladyshev, I. R. Arkhipova, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9352 (2007).
- D. B. Mark Welch, J. L. Mark Welch, M. Meselson, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5145 (2008).
- J. H. Hur, thesis, Harvard University (2006).
- N. N. Pouchkina-Stantcheva, A. Tunncliffe, *Mol. Biol. Evol.* **22**, 1482 (2005).
- See supporting material on Science Online.
- D. M. Baird, J. Rowson, D. Wynford-Thomas, D. Kipling, *Nat. Genet.* **33**, 203 (2003).
- R. Finking, M. A. Marahiel, *Annu. Rev. Microbiol.* **58**, 453 (2004).
- D. Mark Welch, M. Meselson, *Science* **288**, 1211 (2000).
- N. Kondo, N. Nikoh, N. Ijichi, M. Shimada, T. Fukatsu, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14280 (2002).
- K. Suga, D. Mark Welch, Y. Tanaka, Y. Sakakura, A. Hagiwara, *PLoS ONE* **2**, e671 (2007).
- A. Haviv-Chesner, Y. Kobayashi, A. Gabriel, M. Kupiec, *Nucleic Acids Res.* **35**, 5192 (2007).
- T. de Lange, *Genes Dev.* **19**, 2100 (2005).
- R. Jain, M. C. Rivera, J. A. Lake, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3801 (1999).
- C. W. Dunn *et al.*, *Nature* **452**, 745 (2008).
- E. Gladyshev, M. Meselson, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5139 (2008).
- S. P. Otto, T. Lenormand, *Nat. Rev. Genet.* **3**, 252 (2002).
- We thank W. Reznikoff, M. Belfort, and D. Mark Welch for comments and J. Mark Welch, K. van Doninck, and J. Hur for communicating results before publication. Supported by NSF grant MCB-0614142 (M.M. and I.R.A.) and NIH grant GM072708 (M.M.). Sequences obtained in this study were deposited in GenBank (accession numbers EU643473 to EU643504).

Supporting Online Material

www.sciencemag.org/cgi/content/full/320/5880/1210/DC1

Materials and Methods

Figs. S1 to S4

Tables S1 to S12

References

12 February 2008; accepted 22 April 2008

10.1126/science.1156407

Ancestral Monogamy Shows Kin Selection Is Key to the Evolution of Eusociality

William O. H. Hughes,^{1*} Benjamin P. Oldroyd,² Madeleine Beekman,² Francis L. W. Ratnieks³

Close relatedness has long been considered crucial to the evolution of eusociality. However, it has recently been suggested that close relatedness may be a consequence, rather than a cause, of eusociality. We tested this idea with a comparative analysis of female mating frequencies in 267 species of eusocial bees, wasps, and ants. We found that mating with a single male, which maximizes relatedness, is ancestral for all eight independent eusocial lineages that we investigated. Mating with multiple males is always derived. Furthermore, we found that high polyandry (>2 effective mates) occurs only in lineages whose workers have lost reproductive totipotency. These results provide the first evidence that monogamy was critical in the evolution of eusociality, strongly supporting the prediction of inclusive fitness theory.

Eusocial behavior, exemplified by social insects, represents one of the pinnacles of sociality and is characterized by individuals altruistically helping to rear siblings rather than their own offspring (1). The established paradigm, based on inclusive fitness (kin selection) theory, is that eusociality evolves be-

cause of a combination of the direct benefits of altruism (i.e., helpers increase the number of individuals reared) and close relatedness between group members, such that the inclusive fitness of helpers exceeds that achievable through a solitary lifestyle (2–5). High relatedness, arising from the delayed dispersal of offspring [possibly